

化探背景与异常识别的问题与对策

陈明

(中国地质大学·北京·100083)

李金春

(山东地矿厅物探队·济南·250013)

化探背景与异常划分涉及系统误差和不同地质体的背景差异,是地球化学找矿中至关重要的内容,直接影响能否正确提取找矿信息和化探找矿效果。背景与异常划分方法可分成估值和模式识别两大类,无论使用何种方法,都需对原始数据作符合地质规律的转换。

关键词 地球化学找矿 背景与异常划分 C型转换 模式识别



技术·方法

1 背景与异常划分问题由来

自本世纪30年代初前苏联首次做岩石测量,化探背景与异常划分就产生了,但研究地壳的元素地球化学背景早已开始。1924年和1932年, F. W. Clake 两次公布了化学元素在地壳中的平均含量(克拉克值)。其后, A. И. Виноградов (1949)、K. Rankama (1959)、黎彤(1963)和 S. R. Taylor (1994)发表了各自的研究成果。克拉克值现已成为衡量地壳中不同地段和岩石(包括矿石)类型中元素的集中、分散的标准,也用来研究矿源层和特殊地质体的成矿专属性等^[1]。

1963年, H. И. Софронов 提出了“分散晕”的概念,论及元素的分布型式时认为微量元素的含量服从正态分布,为根据正态分布划分背景与异常的各种方法奠定了理论基础。1937年, B. Н. Вернадский 又提出岩体中的“浓度克拉克值”概念,成为衬度类计算方法和在衬度值基础上划分背景与异常的方法的鼻祖。

化探方法开创之始,正值寻找地表矿和近地表矿的鼎盛时期。与地表矿和近地表矿相关的化探样品中指示元素意义明确,元素含量也较高,在原生晕找矿中使用浓度克拉克值以及在土壤测量和水系沉积物测量中使用 $\bar{X} + K \times S$ 方法划分背景与异常虽然在理论上有些牵强之处,但在实际生产中是可行的。这些方法通常在整个研究区内使用统一的背景值,称为全局估计。随着寻找隐伏矿床的需要越来越迫切,全局估计法的不足就暴露出来了。1959年, W. C. Krubein 提出用趋势分析的趋势值表示背景起伏,用剩余值反映异常的空间分布,把背景与异常识别从全局估计引向“局部估计”(常用的移动平均法

也是一种趋势分析法)。1966年, D. G. Krige 提出了一种无偏局部估计法——二维加权移动趋势分析法。这种方法后来发展成为一门独立的学科——地质统计学。化探人员主要使用泛克里格法来划分背景与异常。泛克里格法主要解决了两个问题:1)强调化探背景是起伏的,而且强调化探样品之间的相关性是有一定空间限制的;2)在该限制空间范围内,对漂移值和剩余值的估计在数值计算上是无偏的。

为解决背景与异常划分问题,化探工作者还在化探样品分析手段和背景值估计方面作了努力,如偏提取技术、指示元素赋存状态分析和稳健统计学方法的使用等,这些措施取得了较好的效果。尽管如此,如下问题仍然存在:1)受采样、分析测试和地质成因等方面影响,“背景”不仅不是平面的,也不是一个光滑曲面,而是有一定变化范围的带(图1)。

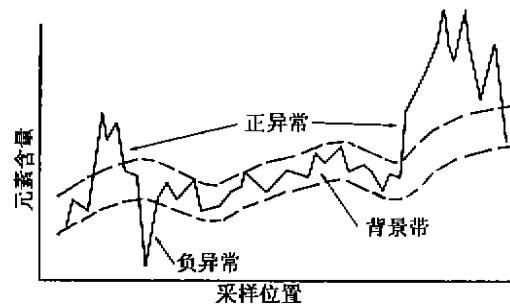


图1 化探背景与异常的关系

因此,剩余值不是异常值,在求得剩余值后合理确定剩余异常下限至关重要;2)在绝大多数情况下,化探样品来自多个不同分布型式和变化规律的地球化学总体。受不同矿化类型和矿化程度地质体剥蚀深度控制,每个总体含量水平和变化范围各不相同。指示隐伏矿体的异常总体,其元素含量可以小于非矿化地质体中该成矿元素含量。这是多重总体分解的地质依据,但分解技术除了常用的对数正态总体分

解法外,近年来没有任何进展。令人遗憾的是,对数正态总体分解方法并没有考虑上述含量变化的实际情况;3)不规范化探数据的系统误差校正方法没有取得实质性突破。不同介质、不同采样与分析批次、不同测试方法和仪器之间的系统误差,与不同总体之间的差异交织在一起,构成划分化探背景与异常及实施其他信息技术之前必须预先解决的难点。

2 若干方法评述

一般把系统误差校正和背景与异常划分看成是两回事。实际上,它们是相互关联的。如图2所示,不失一般性。假定在图2(a)中,图幅1和图幅2是两个存在采样和分析系统误差的相邻图幅,每个图幅内部不存在系统误差,但存在地质体的差异;图2(b)是这两个图幅中的某一化探剖面。由剖面可见3个明显的台阶,这些台阶在数字特征上相同。

消除系统误差的标准方法是使用标准样^[2],这已较圆满地得到解决。但在实践中,仍存在许多不规范的化探工作。当采样与分析测试都是由不同生产单位承担,而且某些单位没有按区域化探规范操作时,标样法就无法实施。

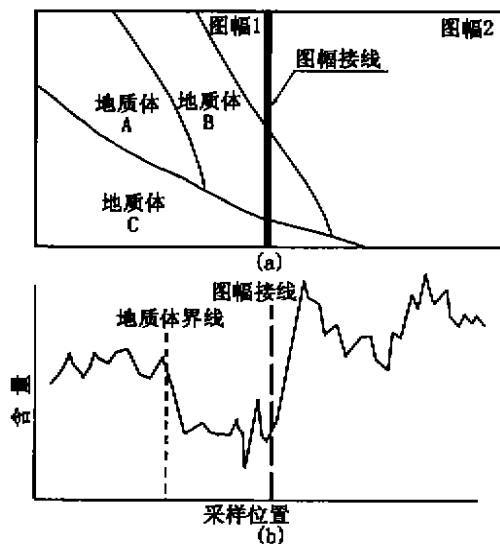


图2 不同图幅的系统误差和不同地质体之间的化探数据的差异

把规范的化探数据与未经校正的非规范数据一起使用只能得出错误的结论,而舍弃非规范数据又太可惜。为此,人们设计和使用了诸如衬度系数法、分区标准化法、移动变差系数法、泛克里格法、多元分析方法等。设计和使用这些方法无非出于以下考虑:要么用一定的含量水平为基准统一数据;要么用一定的地质意义来统一。良好的愿望要用有效的措施来实现,正视这些方法的优缺点对提高化探找矿

效果是极为有益的。

2.1 衬度系数法

衬度系数法操作简单。设 X_{ij} 表示某元素在第 i 幅图内的第 j 个采样点上的分析值, $i = 1, 2, \dots, n$; $j = 1, 2, \dots, m$, 则衬度系数法的具体操作过程可分如下两步操作。第一步:求取该元素在 i 幅图中的平均值 $\bar{X}_i = (\sum_{j=1}^m X_{ij}) / m$; 第二步:计算每个采样点上该元素的衬度值: $X'_{ij} = X_{ij} / \bar{X}_i$ 。

现在我们来考察经衬度系数转换后的数据特征。可以发现,衬度系数法的实质是把各个图幅各个元素的平均含量都统一到数值1,即

$$\bar{X}'_i = \frac{1}{m} \sum_{j=1}^m X'_{ij} = \frac{1}{m} \sum_{j=1}^m \frac{X_{ij}}{\bar{X}_i} = \frac{1}{m} \sum_{j=1}^m \frac{X_{ij}}{\frac{1}{m} \sum_{k=1}^m X_{ik}} = \frac{\sum_{j=1}^m X_{ij}}{\sum_{k=1}^m X_{ik}} = 1 \quad (1)$$

这可以使拼图时图面优化,问题是任意分割的两个不同地区(图幅)的元素平均含量都相等吗?答案是否定的。

一些化探人员为了消除因地质体造成的背景差异,对衬度系数法进行了改进。设 X_{ij}^k 为元素在第 i 个图幅中第 k 类地质体第 j 个采样位置上的元素含量 ($i = 1, 2, \dots, n$; $j = 1, 2, \dots, m_k$, $k = 1, 2, \dots, G$), 其中, n 为总图幅数, m_k 为第 k 类地质体中的样品个数, G 为地质体类型总数,则有

$$\bar{X}_i^k = \frac{1}{m_k} \sum_{j=1}^{m_k} X_{ij}^k \quad (2)$$

$$X'_{ij}^k = \frac{X_{ij}^k}{\bar{X}_i^k} \quad (3)$$

这样处理比原衬度系数法要合理一些,它同时作了系统误差和不同地质体中化探背景差异的校正。但仍然摆脱不了各分幅元素均值都等于1的阴影。

2.2 分区标准化法

也有两种做法。一是在不同图幅之内对分析数据分别进行标准化: $X'_{ij} = (X_{ij} - \bar{X}_i) / S_i$ 。其中 $\bar{X}_i$

含义同上, $S_i = \sqrt{\frac{1}{m_i - 1} \sum_{j=1}^{m_i} (X_{ij} - \bar{X}_i)^2}$ 。另一种做法是,标准化过程不是在某一个图幅内进行,而是通过一个移动窗口,在某一个图幅内进行移动标准化(类似于移动平均法),然后把标准化数据合在一起拼图。

这两种做法都值得深究。经第一种方法处理

后,每个图幅元素含量的平均值都为0,方差均为1,这对化探工作者来说同样不能接受。在第二种做法中,出于总体分解的需要,移动窗口大小必定要与研究区内各地质体的平均尺寸相当。

现假设有两个总体A和B,分别代表某个异常总体和某个背景总体(图3(a))。由于是移动窗口,这种假设在实际工作中是成立的。实际见到的是背景总体与异常总体的混合总体(图3(b))。进行标准化时,相当于把这个混合总体从中间割裂成两部分,标准化后将分别形成如图3(c)所示的均值为0、方差为1的总体。当若干个这样的总体又合到一起时,混合总体的均值和方差仍为0和1,当再用某一异常下限去划分背景与异常时,必然把绝大部分异常总体划归为背景总体,而把一小部分背景总体划归为异常总体。

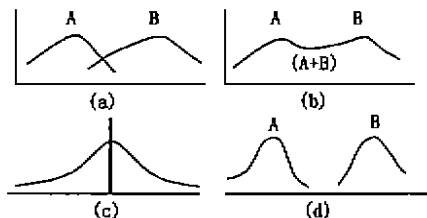


图3 背景总体与异常总体

如此划分的结果肯定是无法认可的。但分区标准化法容易给人造成假象:它能识别弱异常,并把高背景从异常样品中清除出去了。但实际上,被分区标准化法“识别”出来的所谓的弱异常,很可能是背景总体中含量较高者,而被“清除”出来的高背景样品则是真正的异常样品。从分区标准化法的结果不难看出,其缺憾在于它把背景总体和异常总体的差异缩小了,而我们期望的是要把背景总体和异常总体的差异扩大(图3(d))。

可见,系统误差和其他误差校正不是一个简单的拼图问题,其结果将对背景与异常划分,乃至化探找矿效果产生深远的影响,绝不可轻视。

2.3 移动变差系数法

移动变差系数法来源于化探数据的实际表现形式。把矿区与一般无矿区比较,前者具有元素含量变化幅度大的特点。该方法成功之处在于它明确提出化探数据处理方法要以化探找矿原理出发来设计的观点。

在操作步骤方面,移动变差系数法类似于移动平均法,并利用局部变差系数值在一定置信水平下截取若干样品代表异常总体。移动变差系数法尚有不足之处,详见下文。

2.4 泛克里格法

许多生产单位使用泛克里格法来划分背景与异常,用漂移代表背景场,用剩余值来圈定异常。周蒂^[3]建议,在测试以某种自然赋存状态存在的元素的基础上,用泛克里格法求取剩余值,并用剩余值的均方差作为剩余异常下限值来圈定异常。

由于泛克里格法对漂移和剩余值的估计都是无偏的,从数学方法上说无可挑剔。但当用泛克里格法来处理常规化探数据时又出现了新问题:一是没有充分利用化探资料中的多变量信息;二是在原始数据不作预校正的情况下圈定异常,其可信度较差。这正是 Journel 等^[6]特别强调样品“正则化”重要性的原因。目前,克里格法又有了新进展,发展了用于多变量无偏估计的协同克里格法,这对解决第一个问题十分有益。

2.5 多元分析方法

纪宏金^[4]针对化探数据多变量的特点,提出用多变量的马氏距离识别多元总体的异点(离群点)原理。与其他方法一样,它对原始数据的系统误差和背景差异很敏感。

上述5种方法具有代表性,其他相似的做法可依此比较和类推。某些系统误差校正和化探异常识别方法是基于纯数学考虑的,与化探找矿实践有相当距离,在此不予讨论。

3 正确划分化探背景与异常的关键

纵观化探背景与异常划分方法,可以得出如下7点结论:1)在进行背景与异常识别之前,必须选择合理的方法校正采样和分析测试带来的系统误差;2)背景场不是光滑的平面,而是有起伏变化曲面,甚至是有一定厚度的带状曲面;3)在能够正确划分背景与异常之前,往往需要采用一定的手段把背景的起伏变化扣除,把背景场转化成平面的或具有明确的地质意义;4)用于圈定异常的剩余值必须已经消除了系统误差和背景差异;5)要绝对避免把背景总体和异常总体的差异缩小的做法,任何正确的方法都必须把这种差异扩大化;6)拼图和背景与异常划分不是纯数学的,应从化探找矿的实际出发设计相应的方法,处理结果必须便于地质解释;7)尽量利用多变量化探资料消除各种不确定性。

4 单元素系统误差校正及背景与异常划分

单元素系统误差和背景差异校正是多元素信息处理的前提。要校正系统误差并消除背景差异,关

键在于确定一个具有地质依据的、并与地球化学找矿原理和实践相吻合的标准。校正的目的不是让某个特征参数等于某个固定值,而是要使所有的数据具有统一的地质和找矿意义。在图 2(a)中,如果只看到其中一个地球化学图,虽然该图幅与相邻图幅之间系统误差未经校正,有经验的化探专家仍可以通过分析得出合理的关于该图幅内矿产情况的结论。也就是说,在化探专家脑海中存在一种无形的标准,可以用来识别有意义的异常。总结这种标准并把它转化为数学方法,对校正系统误差和消除背景差异无疑是十分有益的。

实践表明,埋藏较浅的矿致异常具有明显的特点,即在一定范围内成矿元素含量高,变化幅度大;埋藏较深的矿致异常,其指示元素也有类似的特点。基于上述考虑,陈明等^[5]曾提出一种数据转换方法—C 型转换方法。C 型转换的计算公式为:

$$C_{ij} = C_{ij}^1 \cdot C_{ij}^2 = \frac{X_{ij}}{\bar{X}_j} \cdot \frac{S_j}{\bar{X}_i} \quad (4)$$

其中, $\bar{X}_j$ 和 S_j 分别是在第 i 个图幅第 j 个采样点附近的某个邻域内元素的平均含量和均方差, C_{ij}^1 为该邻域内的衬度系数, C_{ij}^2 为邻域内相对于整个图幅而言的变差系数。对于背景差异, C 型转换通过在一定的邻域计算 $\bar{X}_j$ 和 S_j 来消除。原作者通过有条件模拟, 确认 C 型转换可把异常和背景总体的差异扩大, 使背景总体向坐标原点靠拢, 异常总体向数轴正方向发散, 这为进一步划分背景与异常提供了良好条件(图 4)。

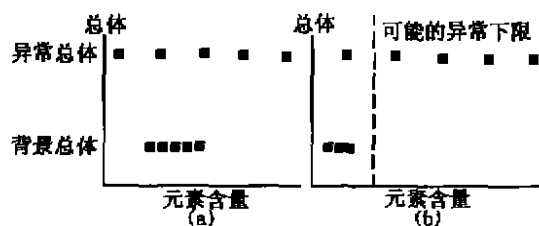


图 4 C 型转换的效果

(a) 转换之前; (b) C 型转换之后

C 型转换的另一个优点是: 在二阶平稳条件下, C 型转换不影响元素之间的相互关系, 因此不影响多元统计分析。

证明: 在二阶平稳条件下, 任意区域化变量有 $E\{X_j\} = m_j$ 和 $\text{Var}\{X_j\} = \delta_j$, 对任意 j , 则

$$\rho_{X'}(X') = \frac{\sum_{j=1}^n (X'_j - \bar{X}'_k)(X'_j - \bar{X}'_l)}{\sqrt{\sum_{j=1}^n (X'_j - \bar{X}'_k)^2 \cdot \sum_{j=1}^n (X'_j - \bar{X}'_l)^2}} =$$

$$\begin{aligned} & \frac{\sum_{j=1}^n \left(\frac{\delta_k \cdot X_{kj}}{m_k^2} - \frac{1}{n} \sum_{i=1}^n \frac{\delta_k \cdot X_{ki}}{m_k^2} \right) \left(\frac{\delta_l \cdot X_{lj}}{m_l^2} - \frac{1}{n} \sum_{i=1}^n \frac{\delta_l \cdot X_{li}}{m_l^2} \right)}{\sqrt{\sum_{j=1}^n \left(\frac{\delta_k \cdot X_{kj}}{m_k^2} - \frac{1}{n} \sum_{i=1}^n \frac{\delta_k \cdot X_{ki}}{m_k^2} \right)^2 \cdot \sum_{j=1}^n \left(\frac{\delta_l \cdot X_{lj}}{m_l^2} - \frac{1}{n} \sum_{i=1}^n \frac{\delta_l \cdot X_{li}}{m_l^2} \right)^2}} \\ & = \frac{\sum_{j=1}^n (X_{kj} - \bar{X}_k)(X_{lj} - \bar{X}_l)}{\sqrt{\sum_{j=1}^n (X_{kj} - \bar{X}_k)^2 \cdot \sum_{j=1}^n (X_{lj} - \bar{X}_l)^2}} = \rho_{Xk}(X) \end{aligned} \quad (5)$$

证毕。

实现 C 型转换的关键在于确定邻域的大小和形状。为此, 需要对每一个相关图幅的各个元素的原始数据进行结构分析, 在确定不同方向上的变程的基础上, 勾划出由变程定义的方向椭圆, 该方向椭圆在统计意义上就是区内各地球化学总体分布空间范围的平均尺寸和形状, 可以作为估值的邻域。C 型转换同时强调了元素的含量水平和变化幅度, 同时具备衬度系数法和移动变差系数法的优点, 既不是仅仅强调其中一个方面, 还避免了它们的缺点。更为重要的是, C 型转换把原始化探数据转化成为一种衡量找矿意义大小的值, 与化探专家考察地球化学图的一般思维方式吻合, 并提出了用找矿意义来统一有系统误差和背景差异的数据的新思路。

该方法已在区域环境地球化学评价^① 和国家 305 项目^② 中得到应用, 效果优于其他方法。例如, 在制作阿尔泰地区 15 个 1:200 000 国际标准分幅的区域地球化学图时遇到困难。这 15 个分幅的区域化探工作有 5 家生产单位分别实施, 个别元素个别图幅的数据是半定量的, 主要成矿元素系统误差在各分幅间十分明显。使用 C 型转换后, 无论用等值线图还是用更为精密的彩色点图显示拼接后的 1:500 000 地球化学图, 各图幅边缘的“台阶”肉眼不可见, 图幅接壤处和地质体界线附近等值线或色区边缘光滑自然。在消除背景差异方面, 因大部分处理结果无法在短期内获得来自勘探方面的验证, 尚不能下绝对的结论。但从利用 C 型转换数据得到的地球化学异常图与已知成矿作用范围、成矿规律、金铜矿产分布等的对比结果看, 具有异常面积较小、异常分布规律与成矿规律吻合程度高、异常与矿产的对应性好等特点。

从化探规律提取化探找矿信息处理技术, C 型转换仅仅是开始, 期望有识之士加强该方面研究。无论如何, 化探数据处理方法不是数字游戏, 任何新

① 陈明, 等, E&G Processor 数据处理系统及其应用, 国家科委“八五”攻关二级课题成果报告(85-925-20-01)。

② 项目编号: (305)02-02。

的方法必须从地质和地球化学找矿实践出发,脱离化探实际的方法是有害而无益的。

5 多重总体分解和多变量背景与异常划分

化探背景与异常划分问题,归根到底是多重总体分解问题。对多重总体分解,一直存在两种不同的作法:一是从数据出发,根据全部所采样品的分析测试值,按某种分布型式划分总体;二是把样品中元素含量与样品的空间分布结合起来,通过模式识别对样品分组归类。这两者都利用了样品的相似性或差异性,但后者还考虑到采样位置,与地球化学找矿原理更加接近。

背景与异常划分的多元方法在实现原理上是多变量统计方法的直接推广,通过多个元素的均值、方差、协方差和多变量空间中的具有一定大小和方向的超椭球结构实现离群点与背景样品的分解。这“使得多元方法具有一元方法无法比拟的功能特点”^[1]。地质统计学方法在处理化探数据方面的最突出优点在于它考虑到了样品之间的空间相关范围,把地质统计学中的结构分析与多重总体分解方法结合起来是目前切实可行的方案。

上述背景与异常划分方法主要是数值估计方法。事实上,多重总体分解的实质是对样品分类,可归结为模式识别问题。多元统计分析提供了判别分析和点群分析方法,分别处理有模型模式识别和无模型模式识别。化探中常常用无模型模式识别来进行总体分解,使用的方法为各种 Q 型聚类分析或自

组织神经网络,其中 Q 型聚类分析在处理大样本数据上有相当的困难,神经网络则比较方便。模式识别是多变量技术,可充分利用化探资料发掘更深层次的找矿信息。但模式识别效果的好坏仍然依赖于原始数据的质量和原始数据的代表性。因此,在模式识别之前对原始数据做合理转换十分重要。

6 小结

用数值估计或模式识别方法进行化探背景与异常划分是地球化学找矿中最基本、最重要的内容之一;能否正确处理这个问题直接关系到能否正确提取找矿信息,直接影响化探找矿效果;系统误差和不同地质体的背景差异是合理划分化探背景与异常的两大障碍,在背景与异常划分或进行其他信息处理之前,必须消除这两者的负面影响;任何化探数据处理方法都必须符合地球化学找矿原理,并具地质和找矿实践依据。

参考文献

- 1 吴锡生.化探数据处理方法.北京:地质出版社,1993
- 2 谢学锦.区域化探.北京:地质出版社,1979
- 3 周蒂.分区背景校正法及其对异常圈定的意义.物探与化探,1984(4)
- 4 纪宏金.地球化学背景与异常划分的多元方法.长春地质学院学报,1983(3)
- 5 陈明,范继璋,矫希国.化探背景与异常划分中的 C 型转换.长春地质学院学报,1996(2)
- 6 Journel A G, Huijbregts C H. Mining Geostatistics. London: Academic Press, 1978

PROBLEM AND COUNTERMEASURE OF RECOGNITION OF GEOCHEMICAL BACKGROUND AND ANOMALY

Chen Ming, Li Jinchun

The systematic error and background differentiation of different geological bodies are concerned in discrimination of background from anomaly, which is an important issue in geochemical prospecting and affects the correct extraction of geochemical information and prospecting effectiveness. The methods of discrimination of anomaly from background are divided into estimating methods and model recognition methods. It is necessary to transform the original data to be consistent with geological regularities.

Key words geochemical exploration, discrimination of anomaly from background, C-transformation, model recognition

第一作者简介:

陈明 男,1967年生。1987年毕业于长春地质学院岩化系,1992年在长春地质学院获硕士学位,1997年获博士学位,目前正在中国地质大学(北京)做博士后研究,现为中国地质大学副教授。主要从事矿产资源预测、金属及油气化探、环境地球化学和数学地质的科研与教学工作。

通讯地址:北京市海淀区学院路29号 中国地质大学(北京)地球化学研究室 邮政编码:100083

